

Gregor Wergen · Jasper Franke ·
Joachim Krug

Correlations between record events in sequences of random variables with a linear trend

Received: date / Accepted: date

Abstract The statistics of records in sequences of independent, identically distributed random variables is a classic subject of study. One of the earliest results concerns the stochastic independence of record events. Recently, records statistics beyond the case of i.i.d. random variables have received much attention, but the question of independence of record events has not been addressed systematically. In this paper, we study this question in detail for the case of independent, non-identically distributed random variables, specifically, for random variables with a linearly moving mean. We find a rich pattern of positive and negative correlations, and show how their asymptotics is determined by the universality classes of extreme value statistics.

Keywords Records · Extreme Value Statistics · Correlations

1 Introduction

The interest in studying record events in a general sense is evident as they by definition tend to be rare events of a special nature. As examples of record-breaking events most readers might think about athletic records like a world record in 100 m dash, or long jump [1,2]. However records also play an important role in other naturally occurring and man-made phenomena. In particular, climatic records attract much attention and are frequently mentioned in the media in the context of global warming. The statistics of record temperatures, precipitation amounts and floods have been addressed extensively in recent work [3,4,5,6,7,8,9,10]. Also in evolutionary biology record events have a certain relevance as was pointed out in [11,12,13], and they

G. Wergen · J. Franke · J. Krug
Institute for Theoretical Physics, University of Cologne
Tel.: +49 221 7396
Fax: +49 221 5159
E-mail: gw@thp.uni-koeln.de, jfranke@thp.uni-koeln.de, krug@thp.uni-koeln.de

play an important role in the dynamics of disordered systems like spin glasses and dirty superconductors [14,15]. Last but not least records can be studied in the financial sciences, e.g. when considering record-breaking values or changes in stock prices [16]. In all these examples not only the rate of record events is of interest but also their correlations: Given that a record flood just occurred, can one relax because it will take a while before the next record flood is to be expected? Based on our previous work in [17], this question will be addressed in the present paper.

Most previous studies of record events and record values [10,18,19,20] have been restricted to the case where the underlying random variables (RV's) $\{X_l\}$ are independently drawn from a common distribution $f(x)$ (i.i.d. RV's). In this case, record events can be shown to be stochastically independent, i.e. the knowledge about previous record events does not indicate whether a future record event is more or less likely. However in many applications, the underlying random variables are not i.i.d. For example, in recent work Majumdar and Ziff computed the statistics of records in a symmetric random walk [21] (see also [16,22,23]). Here, record events are obviously not independent: In a random walk the probability for a second record to occur immediately after a previous record is simply $1/2$, whereas the unconditional probability for a record in the N 'th step is $\approx \frac{1}{\sqrt{\pi N}}$ [21].

In this article we consider correlations between records in sequences of RV's that are independent but not identically distributed. An important example is the statistics of hot or cold records in weather data. In [7] the occurrence of record-breaking temperatures in European and American weather recordings was analyzed by modeling daily temperature measurements as uncorrelated RV's with a linear drift. It was shown that the increase of hot records and the decrease of cold records due to global warming can be described and quantified by this model. We will refer to this particular example of non-i.i.d. RV's, which was originally introduced in the 1980's in the context of athletic records [24,25], as the Linear Drift Model (LDM).

The precise definition of the LDM is as follows: For a positive constant c , each random variable Y_l has an i.i.d. part X_l with density $f(x)$ and an additive linear drift such that

$$Y_l = X_l + cl. \quad (1)$$

In other words, while the shape of the density $f_l(y)$ of the l^{th} RV is invariant, its overall position is shifted by a constant amount c each time a new RV is drawn. As was already noted in [17], the record events under this model are not in general independent (see also [26] for the related case of RV's with changing variance). Below we shall see that the sign and magnitude of correlations depend in a highly non-trivial way on the underlying probability density $f(x)$. To quantify the correlations, consider the joint probability $p_{N,N-1}$ of a record occurring in the $(N-1)^{th}$ as well as in the N^{th} step,

$$p_{N,N-1} \equiv \text{Prob}[Y_N \text{ record and } Y_{N-1} \text{ record}]. \quad (2)$$

To normalize this joint probability, we divide by the probabilities p_N and p_{N-1} of records in the N^{th} and $N-1^{th}$ steps, obtaining the key quantity

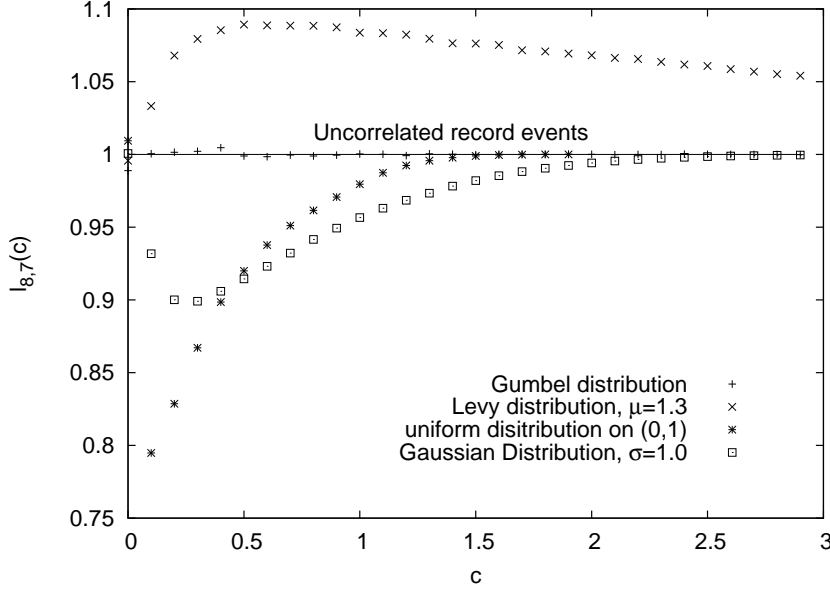


Fig. 1 Correlations between record events in the 7th and 8th RV for different distributions of the underlying i.i.d. part. For the Gumbel case, the well known stochastic independence of record events is found, while simulations for the other distributions show that generally the records are correlated. Note in particular the curve for the symmetric Lévy-stable distribution with index $\mu = 1.3$, which indicates an attraction between record events.

under consideration throughout this paper,

$$l_{N,N-1}(c) \equiv \frac{p_{N,N-1}}{p_N p_{N-1}}. \quad (3)$$

The simulations shown in Fig. 1 illustrate two known results, the fact that record events for i.i.d. RVs are stochastically independent [18,19,20] (all curves meet at unity for $c \equiv 0$), and that for the LDM with Gumbel distributed i.i.d. part, record events are stochastically independent for all values of c [17,19,24,25]. However the curves for other distributions show that stochastic independence does not hold in general. For some choices of the underlying probability density $f(x)$ record events tend to repel each other ($l_{N,N-1}(c) < 1$), whereas in other cases the correlations are positive, i.e. record events attract each other ($l_{N,N-1}(c) > 1$). In particular this last, somewhat counterintuitive numerical finding triggered our interest in this problem.

In the next section we will derive a general expression for the correlations between record events as quantified by $l_{N,N-1}(c)$. For this we employ methods similar to those used in [17]. In Section 3, we provide some explicit examples illustrating the rich and counterintuitive behavior of the correlations. In particular, we discuss the behavior of correlations with respect to the three classes of extreme value statistics [10,27,28,29]. We find that while in the Weibull class of distributions with finite support the correlations are

generally negative, positive correlations mainly arise when the RV's are in the Frechét class of distributions with power law tails. In the intermediate Gumbel class of distributions with exponential-like tails correlations can be both positive and negative. Finally in Section 4, we summarize our results and discuss possible applications and open questions.

2 General theory

Consider a sequence $\{X_l\}$ of independent but not necessarily identically distributed RV's, with $f_l(x)$ denoting the probability density of X_l . The probability that the maximum of the first N RV's is smaller than a given value x then factorizes into the probabilities of the RV's individually being less than x ,

$$\text{Prob}[\max\{X_1, \dots, X_N\} \leq x] = \prod_{l=1}^N \int_{-\infty}^x dx' f_l(x') = \prod_{l=1}^N F_l(x), \quad (4)$$

where $F_l(x) = \int_{-\infty}^x dx' f_l(x')$ denotes the cumulative distribution function for the density $f_l(x)$, and the omitted lower (upper) integral boundary is understood to be the lower (upper) bound of the support of f_l respectively. Throughout we assume distribution functions to be continuous.

Using Eq. 4 the probability p_N that the N^{th} RV is a record can be written as

$$p_N = \int dx f_N(x) \prod_{l=1}^{N-1} F_l(x), \quad (5)$$

and the joint probability that both X_{N-1} and X_N are records is

$$p_{N,N-1} = \int dx_N f_N(x_N) \int_{-\infty}^{x_N} dx_{N-1} f_{N-1}(x_{N-1}) \prod_{l=1}^{N-2} F_l(x_{N-1}). \quad (6)$$

For the LDM we have $f_l(y) = f(y - cl)$ and $F_l(y) = F(y - cl)$, where $f(x)$ and $F(x)$ are the density and distribution function of the i.i.d. part X_l in (1). It then follows that [17]

$$p_N = \int dy f(y - cN) \prod_{l=1}^{N-1} F(y - cl) = \int dx f(x) \prod_{l=1}^{N-1} F(x + cl) \quad (7)$$

and similarly

$$p_{N,N-1} = \int dy f(y) \int_{-\infty}^{y+c} dx f(x) \prod_{l=1}^{N-2} F(x + cl). \quad (8)$$

2.1 Explicit stochastic independence for two cases

The expressions (6) and (8) allow us to verify two known results: Both for the i.i.d. case ($c \equiv 0$) and for the LDM with Gumbel distributed i.i.d. part X_l , record events are independent.

We start with the i.i.d. case. First note that the probability that the N^{th} RV is a record is the probability that it is the largest among N equivalent RV's, so by symmetry $p_N(c=0) = 1/N$. For the i.i.d. case, the subscripts in (6) can be dropped. Using the substitution $u = F(x)$ and $du = f(x)dx$, one obtains

$$p_{N,N-1}(c=0) = \int_0^1 du \int_0^u du' u'^{N-2} = \frac{1}{N-1} \frac{1}{N} = p_N p_{N-1}. \quad (9)$$

It is possible to insert an arbitrary spacing between the two record events and still obtain this factorization to yield $p_{N,N-j} = p_N p_{N-j}$. Similarly, one can consider more than one record event. The factorization property implies that in the i.i.d. case, record events are stochastically independent [18,19,20].

Next we consider the LDM with Gumbel distributed i.i.d. part,

$$f(x) = \exp(-e^{-x} - x) \quad (10)$$

and $F(x) = \exp(-e^{-x})$. Using that in this case [27]

$$F(x+a) = \exp(-e^{-x-a}) = \exp(-e^{-x})^{e^{-a}} = F(x)^{e^{-a}}, \quad (11)$$

the substitution $u = F(x)$ used above yields [17,24]

$$p_N(c) = \int dx f(x) \prod_{l=1}^{N-1} F(x)^{e^{-cl}} = \int_0^1 du u^{\sum_{l=1}^{N-1} \alpha^l} = \frac{1}{\sum_{l=0}^{N-1} \alpha^l} \quad (12)$$

with $\alpha = e^{-c}$, and

$$\begin{aligned} p_{N,N-1}(c) &= \int_0^1 du \int_0^{u^\alpha} du' u'^{\sum_{l=1}^{N-2} \alpha^l} = \frac{1}{\sum_{l=0}^{N-2} \alpha^l} \int_0^1 du u^{\sum_{l=1}^{N-1} \alpha^l} \\ &= \left(\frac{1}{\sum_{l=0}^{N-2} \alpha^l} \right) \left(\frac{1}{\sum_{l=0}^{N-1} \alpha^l} \right) = p_{N-1} p_N. \end{aligned} \quad (13)$$

Again one can introduce an arbitrary number of spaces and intermediate records and still have the corresponding joint probability factorize. This implies that for the LDM with Gumbel distributed i.i.d. part, as for arbitrary distributions without drift, record events are stochastically independent.

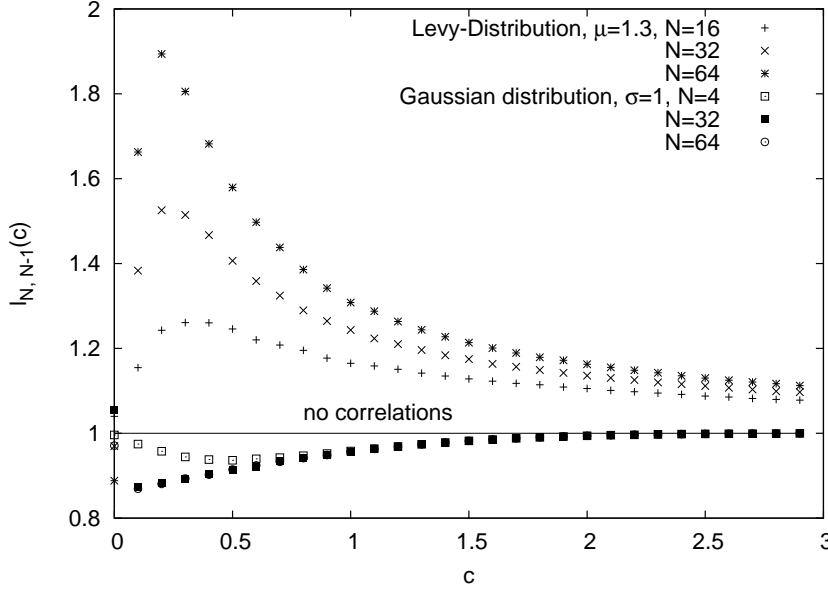


Fig. 2 Correlations between record events as the number of RV's N increases. For the i.i.d. part drawn from a Gaussian distribution with variance $\sigma = 1$, the correlations are negative and for sufficiently large N tend to a limiting form. For the symmetric Lévy distributed i.i.d. part, the correlations are positive and increase with N .

2.2 Expansion for small drift

As shown above, for the i.i.d. case $c \equiv 0$, record events are independent. On the other hand, for very large drift, almost every RV is a record, and thus trivially $l_{N,N-1}(c) \rightarrow 1$ for $c \rightarrow 1$ (see Fig. 1). The behavior of $l_{N,N-1}(c)$ in between these two limits is not as easily characterized. In particular, while for some probability densities the records seem to repel each other for intermediate c ($l_{N,N-1}(c) < 1$), for other probability densities the records seem to *attract* ($l_{N,N-1}(c) > 1$).

Figure 2 shows examples for both types of behavior, illustrating in addition how the correlations evolve with increasing N for the cases of Lévy and Gauss distributed i.i.d. parts. Clearly, the correlations are not a finite size effect, but their N -dependence is markedly different in the two cases. For the Gaussian case, the correlations are negative and seem to approach a form independent of N . For the Lévy case, on the other hand, the correlations seem to increase with N . We will argue later that the correlations eventually become N -independent also for heavy-tailed distributions, but this limit is approached much more slowly than in the Gaussian case.

A first step towards understanding the behavior of the correlations is to determine the sign with which the curves in Figs.1 and 2 depart from the i.i.d. case $c \equiv 0$. To address this question, we compute the record rate $p_N(c)$ and the joint probability $p_{N,N-1}(c)$ in the limit of small c and $c/\sigma \ll N^{-1}$,

where σ is the standard deviation or some other measure of the width of the distribution. For the record rate this problem was discussed extensively in [17] (see also [7]). Knowing that $p_N(c=0) = 1/N$, a Taylor expansion for small c yields

$$p_N(c) \approx \frac{1}{N} + c \frac{N(N-1)}{2} I(N-2), \quad (14)$$

where we have defined the quantity¹

$$I(N) = \int dx f^2(x) F^N(x). \quad (15)$$

In the same small c regime we can also easily determine the joint probability of occurrence of two consecutive records $p_{N,N-1}(c)$. To leading order in c , the expansion of Eq.(8) yields

$$p_{N,N-1}(c) \approx \int dy f(y) \int^y dx f(x) F^{N-2}(x) + c \int dy f^2(y) F^{N-2}(y) \\ + c \frac{(N-1)(N-2)}{2} \int dy f(y) \int^y dx f^2(x) F^{N-3}(x). \quad (16)$$

The first term can be evaluated by partial integration to $\frac{1}{N(N-1)}$, and another partial integration yields

$$p_{N,N-1}(c) \approx \frac{1}{N(N-1)} - c \frac{(N-1)(N-2)-2}{2} I(N-2) \quad (17)$$

$$+ c \frac{(N-1)(N-2)}{2} I(N-3). \quad (18)$$

In addition to the zeroth order term $\frac{1}{N(N-1)}$, we find a positive and a negative correction term that describe the effect of the drift. Depending on which one of these two terms is larger, positive or negative correlations will result.

A similar expansion can be set up for the normalized correlation $l_{N,N-1}(c)$ defined in Eq.(3). Writing

$$l_{N,N-1}(c) = 1 + cJ(N) + \mathcal{O}(c^2), \quad (19)$$

a straightforward computation yields

$$J(N) = N(N-1)I(N-2) + \frac{N(N-1)^2(N-2)}{2} [I(N-3) - I(N-2)] \\ - \frac{N^2(N-1)}{2} I(N-2) - \frac{(N-1)^2(N-2)}{2} I(N-3) \quad (20)$$

Again, both positive and negative terms occur in (20), the relative magnitude of which depends explicitly on the underlying probability density $f(x)$. Thus from (20), it is not at all obvious how the counterintuitive positive

¹ Note that this definition differs slightly from that of the related quantity I_n used in [17]. The integral (15) also appears in the analysis of the density of near-extreme events [30].

correlations observed numerically arise, how their occurrence depends on the properties of f and how the clustering of record events behaves as a function of the number N of RV's considered. Before turning to the explicit evaluation of $J(N)$ in Section 3, we provide some heuristics on the asymptotics of the correlations in the limit of very large N , and address the behavior of correlations between RV's that are more than one time step apart.

2.3 Asymptotics for large N

It is clear from Eqs. (7) and (8) that the asymptotic behavior of the record rate p_N and the joint probability $p_{N,N-1}$ hinges on the existence of the function

$$G_c(x) \equiv \lim_{N \rightarrow \infty} \prod_{j=1}^N F(x + cj) \quad (21)$$

which was rigorously examined by Ballerini and Resnick. In [24], they prove that the limit (21) exists and is nonzero whenever the density $f(x)$ has a finite first moment. For completeness we provide a heuristic version of their argument. Taking the logarithm of Eq. (21) one has to consider the convergence of the series

$$\ln(G_c(x)) = \sum_{j=1}^{\infty} \ln(F(x + cj)). \quad (22)$$

Now $F(x + cj) = \text{Prob}(X \leq x + cj) = 1 - \text{Prob}(X > x + cj)$ and for any finite x and $c > 0$ there is a \tilde{j} large enough such that $\text{Prob}(X > x + c\tilde{j}) \ll 1$. Therefore

$$\ln(G_c(x)) \approx s_{\tilde{j}}(x) - \sum_{j=\tilde{j}}^{\infty} \text{Prob}(X > x + cj), \quad (23)$$

where $s_{\tilde{j}} = \sum_{j=1}^{\tilde{j}-1} \ln(F(x + cj))$ is a finite sum and the logarithms in the remaining series were approximated by their Taylor expansion around unity. The condition for $f(x)$ to have a finite first moment is that $\text{Prob}(X > x)$ decays faster than $1/x$, which implies convergence of the infinite series in (23) and hence of $G_c(x)$ (in the opposite case the series diverges to $-\infty$ and $G_c \equiv 0$). Since $G_c(x)$ is a probability, it is bounded from above by unity and it follows from (7) and (8) that $p_N(c)$ and $p_{N,N-1}(c)$ have finite, non-zero limits for $N \rightarrow \infty$. The same statement then applies to the ratio (3).

Although the focus of this paper is on the case of a positive drift which enhances the occurrence of (upper) records, it is instructive to compare the asymptotic behavior for $c > 0$ described above to the case $c < 0$. For negative drift both the record rate p_N and the joint probability $p_{N,N-1}$ vanish for $N \rightarrow \infty$, and they do so more rapidly than their i.i.d. counterparts. The asymptotic behavior of the ratio (3) is then *a priori* undetermined. We consider a simple example where the quantities of interest can be computed explicitly. Let the probability density of the i.i.d. variables be given by the negative exponential distribution,

$$f(x) = e^x, \quad x \leq 0 \quad (24)$$

and $f(x) = 0$ elsewhere. Then for $c < 0$ the integrand in (7) and (8) is of the form

$$\prod_{l=1}^K F(x + cl) = \exp \left[Kx + \frac{c}{2} K(K+1) \right], \quad x \leq 0, \quad (25)$$

and integration yields

$$p_N(c) = \frac{1}{N} \exp \left[\frac{c}{2} N(N-1) \right], \quad p_{N,N-1}(c) = \frac{1}{N(N-1)} \exp \left[\frac{c}{2} N(N-1) \right]. \quad (26)$$

For $c < 0$ the record rate is suppressed superexponentially, such that the expected number of records remains finite for $N \rightarrow \infty$ (see also [17]). However, forming the ratio (3) we see that

$$l_{N,N-1} = \exp \left[-\frac{c}{2} (N-1)(N-2) \right] \quad (27)$$

which, in contrast to the case of positive drift, grows without bound for $N \rightarrow \infty$. We will see in Section 3 that such a behavior is not restricted to this specific example. The non-existence of the $N \rightarrow \infty$ limit for $c < 0$ suggests that the limiting function

$$l^*(c) \equiv \lim_{N \rightarrow \infty} l_{N,N-1}(c) \quad (28)$$

may not be smooth for $c \rightarrow 0$, as will be explicitly demonstrated in Section 3.

2.4 Dependence on the distance between record events

Before embarking on the study of specific distributions, we briefly comment on the behavior of the correlations as a function of the distance between record events. Denoting by $p_{N,N-k}$ the joint probability for observing a record in the $N-k$ th and the N th event, it was shown in [24] that

$$\lim_{k \rightarrow \infty} \lim_{N \rightarrow \infty} p_{N,N-k}(c) = \left[\lim_{N \rightarrow \infty} p_N(c) \right]^2 \equiv p(c)^2, \quad (29)$$

which implies that record events become uncorrelated for $k \rightarrow \infty$. This leads to the expectation that in a finite time series, the correlations between record events at distance k should decay with k , in agreement with the simulations shown in Fig. 3. The correlations are seen to be maximal at $k = 1$. In the present paper we therefore focus on the case of nearest neighbor correlations, although extending the computations to $k > 1$ is in principle straightforward.

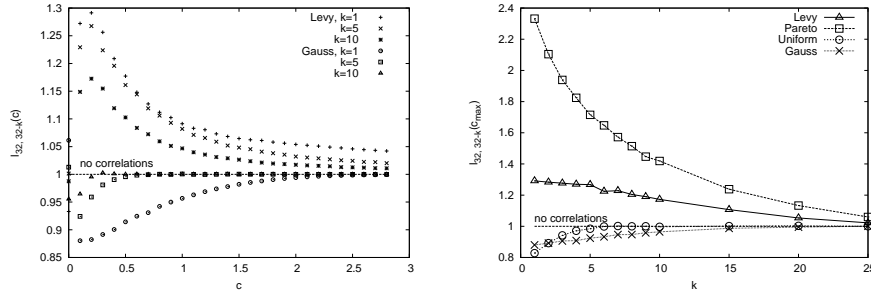


Fig. 3 Correlations between record events at distance k from time series of length $N = 32$ for Gaussian ($\sigma = 1$), Lévy-stable ($\mu = 1.5$), uniform (on $[0, 1]$) and Pareto ($\mu = 1.5$) distributions. **Left:** Full c dependence of $l_{N,N-k}(c)$. **Right:** At the drift velocity c_{\max} where the correlations are maximal ($c_{\max} = 0.1$ for Gaussian and uniform, $c_{\max} = 0.2$ for the Lévy and $c_{\max} = 0.4$ for the Pareto case), $l_{N,N-k}(c_{\max})$ is shown as function of k to illustrate the decay of correlations.

3 Explicit examples

Whereas the statistics of record events in sequences of i.i.d. RV's is completely universal [18, 19, 20], in the presence of drift one has to distinguish between distributions belonging to the three different universality classes of extreme value theory, the classes of Weibull, Gumbel and Fréchet [10, 28, 29]. For each of these classes we will analyze the correlations in the LDM for a few exemplary distributions, starting with the Weibull class, and summarize the observed behavior in a unifying scaling picture in Section 3.4. The results presented below rely on the study of the record rate $p_N(c)$ in the LDM presented in [17], as well as on the expansions for $p_{N,N-1}(c)$ [Eq. (17)] and $l_{N,N-1}(c)$ [Eqs. (19, 20)] derived above in Section 2.2.

3.1 The Weibull class

The Weibull class is the class of distributions with bounded support. As a very simple member of the Weibull class we consider a uniform distribution on the interval $[-a, a]$. We compute $l_{N,N-1}(c)$ in the small c regime by making use of (19, 20) and find for $N \gg 1$

$$l_{N,N-1}(c) \approx 1 - \frac{c}{4a} N^2. \quad (30)$$

The correlations are negative and their magnitude increases rapidly with N (but note that the leading order approximation is valid only for $cN^2/2a \ll 1$). We performed numerical simulations to check this result with $a = 1$ and $-0.1 < c < 0.4$. Figure 4 shows how the descent of $l_{N,N-1}(c)$ for $c > 0$ steepens with increasing N . In the figure we also plot our prediction (30) for small values of N , however the steepness of the descent at zero does not allow us to estimate the quality of our approximation reliably. The figure shows that for large enough N and not too small c , $l_{N,N-1}(c)$ eventually becomes independent of N , as expected from the general arguments given in Section

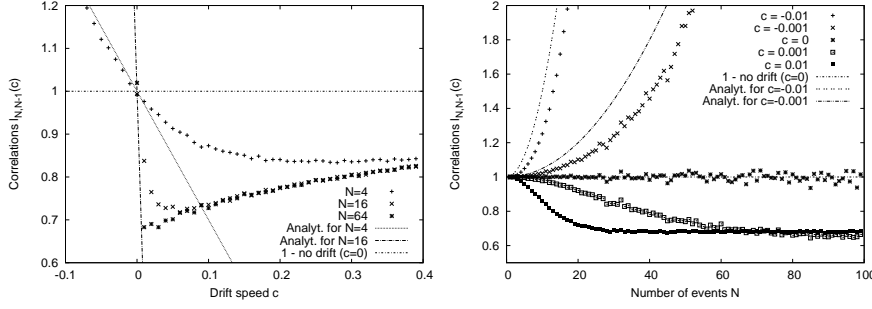


Fig. 4 **Left:** Numerical simulations of $l_{N,N-1}(c)$ for a uniform distribution of width $a = 1$. For each c and N we simulated 10^6 time series. The figure shows data for $N = 4, 16$ and 64 . At $c = 0$ we find a steep descent of $l_{N,N-1}(c)$. For $N = 4$ and $N = 16$ we show the analytic prediction (30) (steep lines). With increasing c , $l_{N,N-1}(c)$ appears to become completely independent of N . **Right:** Simulation results for $l_{N,N-1}$ as a function of N for $c = -0.01, -0.001, 0, 0.001$ and 0.01 . We simulated 10^7 series of 100 RV's in each case. For $c < 0$ we find a very steep increase of $l_{N,N-1}$ that is in good agreement with the analytical prediction (30). For $c > 0$, $l_{N,N-1}$ quickly saturates at a constant value and our analytical results are not very useful.

2.3. In this case the approach to the $N \rightarrow \infty$ limit is particularly rapid, because distributions that are more than $2a/c$ steps apart do not overlap. The sharp initial drop of $l_{N,N-1}(c)$ indicates that the limiting distribution $l^*(c)$ is discontinuous at $c = 0$, which is consistent with the non-existence of a limiting function² for $c < 0$.

As a more general subset of the Weibull class we next consider the class of distributions defined by

$$f_\xi(x) = \xi(1-x)^{\xi-1} \quad (31)$$

with $\xi > 0$ and $0 < x \leq 1$. For $\xi = 1$ this produces a uniform distribution similar to the one used above. In [17] we found that for $\xi > \frac{1}{2}$ the integral in (15) is given by

$$I(N) = \xi \frac{\Gamma\left(2 - \frac{1}{\xi}\right) \Gamma(N+1)}{\Gamma\left(N + 3 - \frac{1}{\xi}\right)}. \quad (32)$$

Using this and assuming $N \gg 1$ we find the following expression for $l_{N,N-1}(c)$:

$$l_{N,N-1}(c) \approx 1 - \frac{cN^3}{2} \frac{\Gamma\left(2 - \frac{1}{\xi}\right) \Gamma(N-1)}{\Gamma\left(N + 1 - \frac{1}{\xi}\right)} \quad (33)$$

² With regard to its upper tail, the negative exponential distribution (24) is equivalent to the uniform distribution, and therefore the argument of Section 2.3 applies here as well.

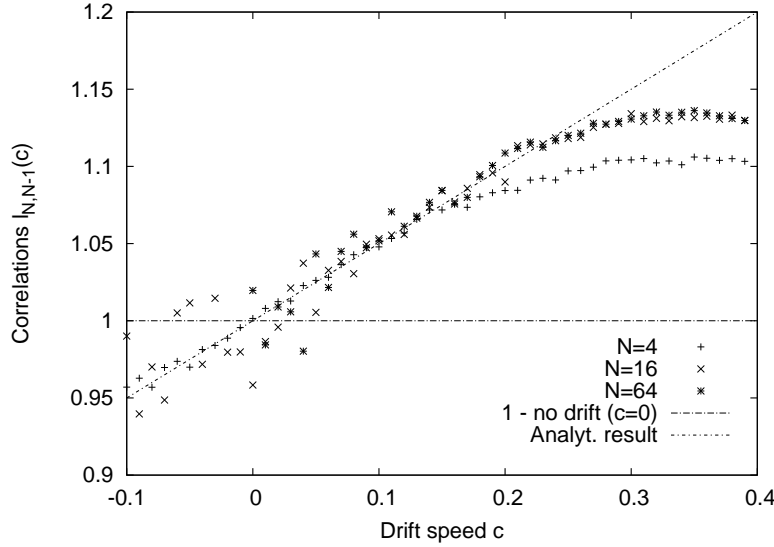


Fig. 5 Numerical Simulations of $l_{N,N-1}(c)$ for an exponential distributions with unit mean. For each c and N we simulated 10^6 time series. The figure shows results for $N = 4, 16$ and 64 along with our N -independent analytical prediction (35).

and making use of the Stirling approximation we finally obtain

$$l_{N,N-1}(c) \approx 1 - \frac{c}{2} \Gamma\left(2 - \frac{1}{\xi}\right) N^{1+\frac{1}{\xi}}. \quad (34)$$

Similar to the uniform case, the correlations between neighboring record events are always negative and their magnitude increases rapidly with increasing N , suggesting a discontinuity of the limiting function $l^*(c)$ at $c = 0$.

3.2 The Gumbel class

As a first example in the Gumbel class we consider the exponential distribution with mean a , $f(x) = a^{-1}e^{-\frac{x}{a}}$, $x \geq 0$. In [17] it was found that in this case the effect of the linear drift on the record rate is independent of N to leading order in c . Using this result it is straightforward to obtain

$$l_{N,N-1}(c) \approx 1 + \frac{c}{2a}. \quad (35)$$

We compared this result to numerical simulations in Fig. 5 and found a very good agreement. Apparently $l_{N,N-1}(c)$ assumes a value independent of N already for very small N .

Another important member of the Gumbel class is the normal distribution with mean zero and standard deviation σ . Here, the computation is usually

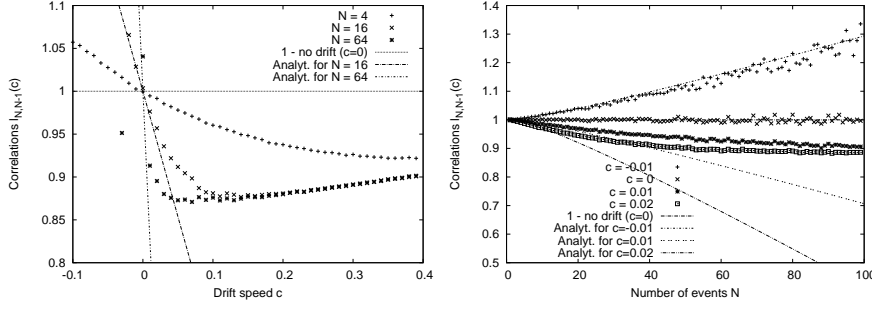


Fig. 6 Numerical simulations of $l_{N,N-1}(c)$ for a Gaussian distribution of width $\sigma = 1$. **Left figure** shows $l_{N,N-1}(c)$ for $N = 4, 16$ and 64 . For each c and N we simulated 10^7 time series. The figure illustrates the steep descent of the correlator at $c = 0$ when c is small. **Right figure** shows $l_{N,N-1}(c)$ for different, fixed values of c together with our analytical results. Here, we analyzed 10^8 series of RV's for each drift rate. Again, we find agreement between the simulations and our analytical computations for small N and c . We manually fitted curves $\propto N\sqrt{\ln(N^2/8\pi)}$. Interestingly for $c < 0$ the agreement is a lot better and $l_{N,N-1}(c)$ does not saturate at a constant value.

a bit more complicated; however, as in the previous examples, most of the work was already done in [17]. There we found that for $N \gg 1$ and $cN \ll \sigma$

$$I(N) \approx \frac{c}{N^2\sigma} \frac{4\sqrt{\pi}}{e^2} \sqrt{\ln\left(\frac{N^2}{8\pi}\right)}. \quad (36)$$

Using (20) this result allows us to estimate the effect of the drift on the correlations. We find

$$(l_{N,N-1}(c) - 1) \propto -N \frac{c}{\sigma} \sqrt{\ln\left(\frac{N^2}{8\pi}\right)}. \quad (37)$$

Unlike the case of the exponential distribution, in the Gaussian case the correlations are negative and depend strongly on N . The behavior is similar to that found in the Weibull case, and matches the result (34) (up to logarithmic factors) for $\xi \rightarrow \infty$. Numerical results for the Gaussian distribution are shown in Fig. 6. Similar to the uniform distribution, $l_{N,N-1}(c)$ shows a steep descent at $c = 0$ which becomes steeper for increasing N . However, the descent appears to be smoother than in the uniform case. In agreement with the considerations of Section 2.3, for $c > 0$ and large N the correlations become independent of N . In contrast, for $c < 0$ the correlations increase monotonically with N and do not appear to saturate. In the case of the Gaussian distribution our analytical prediction (37) is only valid in a small regime for small $c > 0$ and relatively small N , but for $c < 0$ the approximation is significantly better.

To understand the marked difference between the exponential and Gaussian distributions, we analyze the general class of Gumbel-type distributions of the form

$$f(x) = A_\beta e^{-|x|^\beta}, \quad (38)$$

with normalization A_β and $\beta > 0$. In this case it was found in [17] that

$$I(N) \propto N^{-2} \ln(N)^{1-\beta^{-1}}. \quad (39)$$

With some further computations this leads us to the following behavior of the leading order correction coefficient $J(N)$:

$$J(N) \propto -D_1 \left(1 - \frac{1}{\beta}\right) N \ln(N)^{1-\frac{1}{\beta}} + D_2 \ln(N)^{1-\frac{1}{\beta}}, \quad (40)$$

where D_1 and D_2 are positive constants not depending on N . The exact values of D_1 and D_2 are very difficult to compute and we will not consider them in this article. Nevertheless, this expression nicely reproduces our results for the exponential and the Gaussian distribution and shows how both positive and negative correlations may emerge in the Gumbel class. For $\beta = 1$ the leading first term vanishes and $J(N)$ reduces to a positive constant. For all $\beta \neq 1$ the first term dominates and the correlations grow with N , with a sign determined by $1 - \frac{1}{\beta}$. For values $\beta > 1$ of distributions that decay faster than the standard exponential we find negative correlations between the records, in agreement with the Gaussian example, while for stretched exponential distributions ($\beta < 1$) positive, N -dependent correlations result (see Fig. 8 for an example).

In hindsight, the special role of the exponential distribution should not come as a surprise, since the Gumbel distribution (10) (for which correlations are completely absent) has an exponential tail. In the sense of extreme value theory, the exponential distribution is close to the Gumbel, and the corresponding records in the LDM are therefore almost uncorrelated (up to small residual correlations which are independent of N).

3.3 The Fréchet class

As a subset of the distributions in the Fréchet class of extreme value statistics we consider the well known Pareto distribution

$$f(x) = \mu x^{-\mu-1} \quad (41)$$

with $\mu > 1$ and $x \geq 1$. In [17] it was found that in this case

$$I(N) = \mu \frac{\Gamma\left(2 + \frac{1}{\mu}\right) \Gamma(N+1)}{\Gamma\left(N + 3 + \frac{1}{\mu}\right)}. \quad (42)$$

This allows us to compute $l_{N,N-1}(c)$ for the Pareto distribution. We find

$$J(N) = \frac{\mu}{2} \frac{\Gamma\left(2 + \frac{1}{\mu}\right) \Gamma(N-1)}{\Gamma\left(N + \frac{1}{\mu} + 1\right)} \frac{N(N-1)(N^2 + \mu N + \mu)}{\mu N + \mu + 1} \quad (43)$$

which leads to the expression

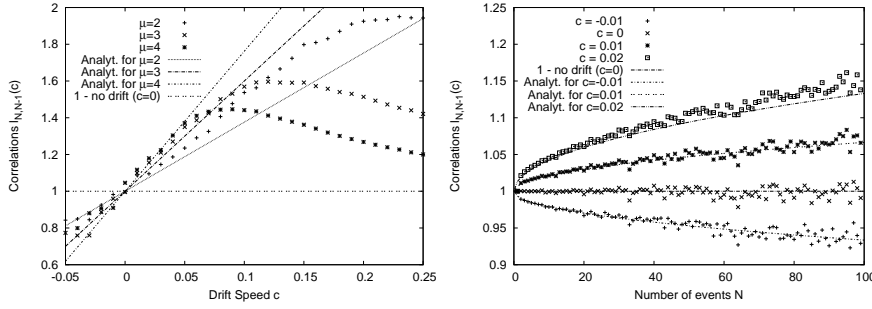


Fig. 7 Left: Numerical simulations of $l_{N,N-1}(c)$ for Pareto distributions. For each c and μ we simulated 10^6 time series of length $N = 16$. The figure shows simulations for $\mu = 2, 3$ and 4 , along with our analytical predictions for small c . At $c = 0$ we find a steep ascent of $l_{N,N-1}(c)$ which gets steeper with increasing μ . Our analytical result from (44) is in good agreement with the simulations for small enough c . **Right:** Simulation results for a Pareto distribution with $\mu = 2$ and different values of c . We analyzed 10^8 series of $N = 100$ RV's for each c . Our analytical work predicts a square-root behavior of $l_{N,N-1}(c) - 1$, which is in good agreement with the simulations.

$$l_{N,N-1}(c) \approx 1 + \frac{c}{2} \Gamma\left(2 + \frac{1}{\mu}\right) N^{1-\mu^{-1}} \quad (44)$$

for large N . For positive c , we thus expect positive correlations between neighbouring record events for all distributions of Pareto form. These correlations are increasing with N slower than linearly. We expect that, asymptotically, this behavior is universal for all distributions within the Fréchet class³. We compare Eq.(44) to numerical simulations in Fig. 7, finding good agreement for small c , both for the c - and the N -dependence. The agreement improves when the distribution becomes broader for smaller μ . The strength of correlations for strongly heavy-tailed distributions of the Pareto class is remarkable. For example, for a Pareto distribution with coefficient $\mu = 2$ we found correlations $l_{N,N-1} \approx 3$ for large N (see Fig. 8).

The correlations displayed in the right panel of Fig.7 increase (for $c > 0$) or decrease (for $c < 0$) with N without showing any sign of saturation, although we know from the general considerations of Section 2.3 that $l_{N,N-1}$ must approach an N -independent limit for $c > 0$ and $\mu > 1$. As we will see in the next subsection, this reflects the fact that the time scale N^* at which the limit is attained is (for a given value of c) particularly large for distributions of the Fréchet class. We also note that because $l_{N,N-1} < 1$ for $c < 0$, the normalized correlations are confined to lie between 0 and 1 in this case. This implies that the divergence of $l_{N,N-1}(c)$ for $N \rightarrow \infty$ and $c < 0$, which was demonstrated in Section 2.3 to occur for a specific example in the Weibull class, cannot happen here. We therefore conjecture that, for distributions in the Fréchet class, the limit (28) exists also for $c < 0$.

³ In [17] we presented data for a Lévy stable distribution with $\mu = 1.8$ which show negative correlations between record events. We have checked that this behavior is not asymptotic, and that the correlations become positive for larger values of N .

3.4 Unified picture

The results of the preceding subsections can be summarized in a simple scaling picture. We first recall from [17] and [22] that the LDM (with $c > 0$) contains a characteristic time scale N^* at which the record rate $p_N(c)$ crosses over from the i.i.d. behavior $p_N \approx \frac{1}{N}$ to the limiting value $p(c) = \lim_{N \rightarrow \infty} p_N(c) > 0$. For the different distributions discussed above, this time scale diverges for $c \rightarrow 0$ according to

$$N^* \propto c^{-\frac{\xi}{1+\xi}} \quad \text{Weibull class} \quad (45)$$

$$N^* \propto c^{-1} |\ln c|^{\frac{1}{\beta}-1} \quad \text{Gumbel class} \quad (46)$$

$$N^* \propto c^{-\frac{\mu}{\mu-1}} \quad \text{Fréchet class } (\mu > 1). \quad (47)$$

Ignoring for the moment the logarithmic factor in the Gumbel case, these behaviors can be further simplified by expressing the different universality classes in terms of the generalized Pareto distribution [31]

$$f(x) = (1 + \kappa x)^{-\frac{\kappa+1}{\kappa}}. \quad (48)$$

For $\kappa > 0$ this is of Pareto type with $\mu = \frac{1}{\kappa}$, for $\kappa < 0$ it reduces to a Weibull-type distribution similar to (31) with $\xi = -\frac{1}{\kappa}$, and the Gumbel class is represented by the exponential distribution which arises from (48) for $\kappa \rightarrow 0$. Using this representation, the different cases in (45), (46) and (47) reduce to

$$N^* \propto c^{-\nu} \quad \text{with } \nu = \frac{1}{1-\kappa}. \quad (49)$$

Note that $\nu < 1$ in the Weibull class but $\nu > 1$ in the Fréchet class, which explains the slow convergence to the $N \rightarrow \infty$ limit in the latter case. Moreover, it was shown in [17] that the integral (15) behaves for large N as (see also [30])

$$I(N) \sim N^{\frac{1}{\nu}-3} = N^{-(2+\kappa)}. \quad (50)$$

To relate this to the behavior of the correlations, we note that for large N Eq.(20) can be approximately written as

$$J(N) \approx -\frac{1}{2} N^4 \frac{d}{dN} I(N) - N^3 I(N) + \mathcal{O}(N^2 I(N)). \quad (51)$$

Inserting (50) we thus conclude that, to leading order,

$$J(N) \approx \frac{\kappa}{2} N^3 I(N) \sim N^{1-\kappa}, \quad (52)$$

which correctly reproduces both the sign and the order of magnitude of the correlations derived in Section 3 for the Weibull and Fréchet classes: Correlations are positive (negative) for $\kappa > 0$ ($\kappa < 0$), and they scale sublinearly (superlinearly) with N in the two cases. A similar calculation using the refined estimate (39) of $I(N)$ for the Gumbel class shows that the correlations are negative (positive) for $\beta > 1$ ($\beta < 1$), in agreement with (40).

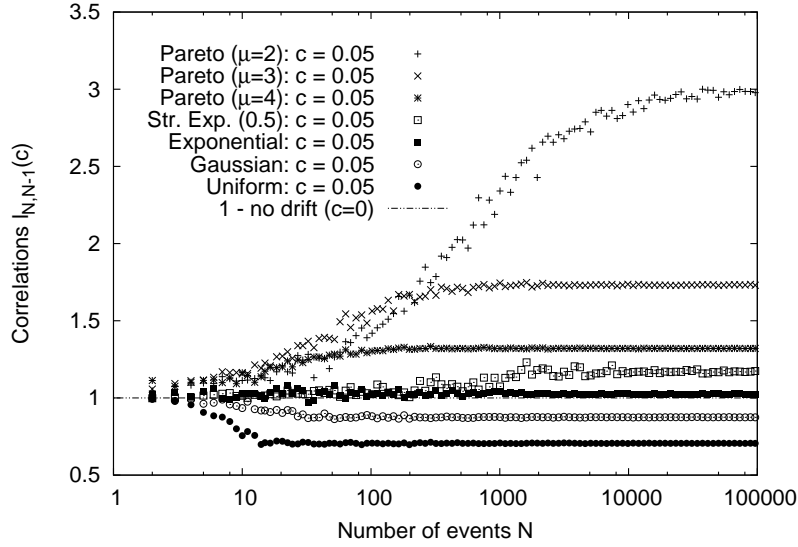


Fig. 8 Numerical simulations of $l_{N,N-1}(c)$ with $c = 0.05$ for six different distributions. In each case we averaged over 10^6 time series of length $N = 10^5$. The results for $l_{N,N-1}(c)$ were binned logarithmically in order to improve the averaging for large N . For the uniform and the Gaussian distribution $l_{N,N-1}(c)$ is clearly negative for $c > 0$ and it decreases with growing N . At some $N = N^*$ the correlations become independent of N . For the special case of the exponential distribution we find a constant value of $l_{N,N-1}(c)$ which is only slightly larger than unity. Both the stretched exponential (with $\beta = 1/2$) and the Pareto distributions (with $\mu = 2, 3, 4$) show positive correlations. Note the slow convergence for the Pareto distribution with $\mu = 2$.

4 Conclusion

In this paper we analyzed the effect of a linear drift on the correlations between records drawn from series of independent RV's, as quantified by the normalized joint probability $l_{N,N-1}(c)$. In Section 2 we derived general expressions (exact and approximate) for $l_{N,N-1}(c)$ and recalled the fact that records are independent both for i.i.d. RV's ($c = 0$) and in the special case where the RV's are drawn from the Gumbel distribution.

Our main analytic results were obtained in Section 3 by way of an expansion in the small c limit, similar to the approach developed previously in [17]. Using this approach we were able to show that the correlations are generally negative ('repulsive') for distributions in the Weibull class, and positive ('attractive') for distributions in the Fréchet class. In the Gumbel class the sign of the correlations depends on the stretching exponent β in (38), with the border between positive and negative correlations being given by the exponential case $\beta = 1$. In contrast to all other cases, for distributions with an exponential tail the correlations are weak and independent of N , which is consistent with the fact that this class of distributions also contains the Gumbel distribution, which has no correlations at all. Simulation results

illustrating the different cases are summarized in Fig. 8. A special role of the exponential distribution in separating two regimes of qualitatively different behaviors has been noted previously in the related context of near-extreme events [30].

Perhaps the most surprising and counterintuitive outcome of our work is the discovery of strong positive record correlations for distributions with a power law tail⁴. In view of the substantial interest in detecting and explaining heavy-tailed distributions in all areas of science [29,32], our finding suggests that drift-induced record correlations could be used as a distribution-free test for detecting power laws or stretched exponentials in empirical data [33].

An interesting open question concerns the structure of the record correlations in the asymptotic limit $N \rightarrow \infty$, where the record rate approaches a nonzero constant and the record process thus becomes stationary. Based on the work of Ballerini and Resnick [24], we have argued in Section 2.3 that the limit (28) exists for $c > 0$ whenever the underlying distribution has a finite mean, whereas for $c < 0$ it is possible that $l_{N,N-1}(c)$ diverges for $N \rightarrow \infty$. The fact that the coefficient of the leading order term in the small c expansion generally diverges with N indicates that the limiting function $l^*(c)$ may be singular for $c \rightarrow 0$, and we have presented numerical evidence for the occurrence of a discontinuity at $c = 0$ for the Weibull class. For heavy-tailed distributions in the Fréchet class, the large- N asymptotics is difficult to ascertain numerically because of the slow convergence, but we have argued that in this case $l^*(c)$ may exist also for $c < 0$. Rigorous work addressing these questions along the lines of [24] would be most welcome.

Acknowledgements This work has been supported by DFG within the *Bonn Cologne Graduate School of Physics and Astronomy*, by Friedrich Ebert Stiftung through a fellowship to GW, and by Studienstiftung des Deutschen Volkes through a fellowship to JF.

References

1. Gembris, D., Taylor, J.G., Suter D.: Sports statistics: Trends and random fluctuations in athletics. *Nature* **417**, 506 (2002)
2. Gembris, D., Taylor, J.G., Suter D.: Evolution of athletic records: Statistical effects versus real improvements. *J. Appl. Stat.* **34**, 529 (2007)
3. Bassett, G.W.: Breaking recent global temperature records. *Climatic Change* **21**, 303 (1992)
4. Benestad, R.E.: How often can we expect a record event? *Climate Research* **25**, 3 (2003)
5. Redner, S., Petersen, M. R.: Role of global warming on the statistics of record-breaking temperatures. *Phys. Rev. E* **74**, 061114 (2006)
6. Meehl, G.A., Tebaldi, C., Walton, G., Easterling, D., McDaniel, L.: Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U.S. *Geophys. Res. Lett.* **36**, L23701 (2009)
7. Wergen, G., Krug, J.: Record-breaking temperatures reveal a warming climate. *EPL* **92**, 30008 (2010)

⁴ We note in this context that in a previous study of records drawn from series of independent RV's with an increasing variance only negative correlations were found [26].

8. Newman, W.I., Malamud, B.D., Turcotte, D.L.: Statistical properties of record-breaking temperatures. *Phys. Rev. E* **82**, 066111 (2010)
9. Anderson, A., Kostinski, A.: Reversible record breaking and variability: Temperature distributions across the globe. *J. Appl. Meteor. Climat.* **49**, 1681 (2010)
10. de Haan, L., Ferreira, A.: *Extreme Value Theory - An Introduction*. Springer, New York (2006)
11. Sibani, P., Brandt, M., Alstrøm, P.: Evolution and extinction dynamics in rugged fitness landscapes. *Int. J. Mod. Phys.* **12**, 361 (1998)
12. Krug, J., Jain, K.: Breaking records in the evolutionary race. *Physica A* **358**, 1 (2005)
13. Park, S.-C., Krug, J.: Evolution in random fitness landscapes: the infinite sites model. *J. Stat. Mech.: Theor. Exp.* P04014 (2008)
14. Oliveira, L.P., Jensen, H.J., Nicodemi, M., Sibani, P.: Record dynamics and the observed temperature plateau in the magnetic creep-rate of type-II superconductors. *Phys. Rev. B* **71**, 104526 (2005)
15. Sibani, P., Rodriguez, G.F., Kenning, G.G.: Intermittent quakes and record dynamics in the thermoremanent magnetization of a spin-glass. *Phys. Rev. B* **74**, 224407 (2006)
16. Wergen, G., Bogner, M., Krug, J.: Record statistics for biased random walks, with an application to financial data. *Phys. Rev. E* **83**, 051109 (2011)
17. Franke, J., Wergen, G., Krug, J.: Records and sequences of records from random variables with a linear trend. *J. Stat. Mech.: Theor. Exp.* P10013 (2010)
18. Arnold, B.C., Balakrishnan, N., Nagaraja, H.N.: *Records*. Wiley, New York (1998)
19. Nevzorov, V. B.: *Records: Mathematical Theory*. American Mathematical Society (2001)
20. Glick, N.: Breaking Records and Breaking Boards. *Amer. Math. Monthly* **85**, 2 (1978)
21. Majumdar, S.N., Ziff, R.M.: Universal Record Statistics of Random Walks and Lévy flights. *Phys. Rev. Lett.* **101**, 050601 (2008)
22. Le Doussal, P., Wiese, K.J.: Driven particle in a random landscape: Disorder correlator, avalanche distribution, and extreme value statistics of records. *Phys. Rev. E* **79**, 051105 (2009)
23. Sabhapandit, S.: Record Statistics of Continuous Time Random Walk. *EPL* **94**, 20003 (2011)
24. Ballerini, R., Resnick, S.: Records from improving populations. *J. Appl. Prob.* **22**, 487-502 (1985)
25. Ballerini, R., Resnick, S.: Records in the presence of a linear trend. *Adv. Appl. Prob.* **19**, 801-828 (1987)
26. Krug, J.: Records in a changing world. *J. Stat. Mech.: Theor. Exp.* P07001 (2007)
27. Gumbel, E. J.: *Statistical Theory of Extreme Values and Some Practical Applications*. National Bureau of Standards Applied Mathematics Series **33** (1954)
28. Galambos, J.: *The Asymptotic Theory of Extreme Order Statistics*. Wiley, New York (1987)
29. Sornette, D.: *Critical Phenomena in Natural Sciences*. Springer, Berlin (2004)
30. Sabhapandit, S., Majumdar, S.N.: Density of near-extreme events. *Phys. Rev. Lett.* **98**, 140201 (2007)
31. Pickands III, J.: Statistical inference using extreme order statistics. *Ann. Stat.* **3**, 119 (1975)
32. Christensen, K., Moloney, N.R.: *Complexity and Criticality*. Imperial College Press, London (2005)
33. Franke, J., Wergen, G., Krug, J. (in preparation)